HW1.

2022-18599 이기현

#1.

(a) $l_i(\theta) = \frac{1}{2}(x_i^T\theta - Y_i)^2$. Let $x_i^T = [X_{i1}, X_{i2}, \cdots, X_{ip}]$

$$= \frac{1}{2}\left(\sum_{j=1}^{p} X_{ij}\cdot\theta_j - Y_i\right)^2.$$

$$\frac{\partial}{\partial\theta_j} l_i(\theta) = \left(\sum_{j=1}^{p} X_{ij}\cdot\theta_j - Y_i\right)\cdot X_{ij} = (x_i^T\theta - Y_i)\cdot X_{ij}$$

$$\therefore \nabla_\theta l_i(\theta) = (x_i^T\theta - Y_i)\cdot X_i$$

(b) $\mathcal{L}(\theta) = \frac{1}{2}\|X\theta - Y\|^2 = \frac{1}{2}\sum_{i=1}^{N} 2l_i(\theta) = \sum_{i=1}^{N} l_i(\theta)$

$$\therefore \nabla_\theta \mathcal{L}(\theta) = \sum_{i=1}^{N}\nabla_\theta l_i(\theta) = \sum_{i=1}^{N}(x_i^T\theta - Y_i)X_i = \sum_{i=1}^{N}(x_i^T\theta X_i - Y_iX_i)$$

$$= X^TX\theta - X^TY. \quad \left(\because x_i^T\theta X_i = \begin{pmatrix} X_{i1}^2\theta_1 \\ \vdots \\ X_{ip}^2\theta_p \end{pmatrix} = x_i^Tx_i\theta.\right)$$

#2. $\theta^{k+1} = \theta^k - \alpha f'(\theta^k) = \theta^k - \alpha\theta^k = (1-\alpha)\theta^k$.

if $\alpha > 2$, $|\theta^{k+1}| = |1-\alpha|\cdot|\theta^k|$.

By induction, $|\theta^{k+1}| = |1-\alpha|^{k+1}\cdot|\theta^0|$. $|1-\alpha| > 1$.

For $\forall N$, let $k = \left\lceil \log_{|1-\alpha|}\frac{N}{|\theta^0|}\right\rceil$. then

$|\theta^k| = |1-\alpha|^k|\theta^0| \geq \frac{N}{|\theta^0|}\cdot|\theta^0| = N$. hence θ^k diverges.

#3. $\nabla f(\theta^k) = X^T(X\theta^k - Y)$ due to (1b).

$\therefore \theta^{k+1} = \theta^k - \alpha X^T(X\theta^k - Y) = (1 - \alpha X^TX)\theta^k + \alpha X^TY$.

Say $\theta^* = (X^TX)^{-1}X^TY$.

$$\theta^{k+1} - \theta^* = (I - \alpha X^TX)\theta^k + \alpha X^TY - \theta^*$$

$$= (I - \alpha X^TX)(\theta^k - \theta^*) + \alpha X^TY - \alpha X^TX\theta^*$$

$$= (I - \alpha X^T X)(\theta^k - \theta^*) + \cancel{\alpha X^T Y - \alpha X^T X (X^T X)^{-1} X^T Y}$$

Let the eigen values of $X^T X$ be $\lambda_1 \dots \lambda_n$ respectively.

and $0 < \lambda_1 \leq \dots \leq \lambda_n$. $\alpha > \frac{2}{\lambda_n}$

Then. the eigen values of $I - \alpha X^T X$ are $1 - \alpha\lambda_1, \dots, 1 - \alpha\lambda_n$.

as $\alpha > 0$.

Then, we have $1 - \alpha\lambda_n < 1 - \frac{2}{\lambda_n} \cdot \lambda_n = -1$.

Say $D = Q^T (I - \alpha X^T X) Q$ where Q is orthogonal and

$$D = \begin{pmatrix} 1-\alpha\lambda_1 & & \\ & \ddots & \\ & & 1-\alpha\lambda_n \end{pmatrix}$$ diagonal.

As $\theta^k - \theta^* = Q D^k Q^T (\theta^0 - \theta^*)$. we have

$\|\theta^k - \theta^*\| = \|D^k Q^T (\theta^0 - \theta^*)\|$. Denote $y = Q^T(\theta^0 - \theta^*) = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$.

Then $D^k y = \begin{pmatrix} (1-\alpha\lambda_1)^k y_1 \\ \vdots \\ (1-\alpha\lambda_n)^k y_n \end{pmatrix}$

If $y_n \neq 0$, $\|D^k y\| \geq |1 - \alpha\lambda_n|^k |y_n|$, which diverges to ∞.

Hence $\theta^k - \theta^*$ can converge for at most θ^* with $y_n = 0$.

This means $\theta^0 - \theta^* = Q \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = Q \begin{pmatrix} y_1 \\ \vdots \\ y_{n-1} \\ 0 \end{pmatrix}$.

Because Q is invertible, $V = \left\{ Q \begin{pmatrix} y_1 \\ \vdots \\ y_{n-1} \\ 0 \end{pmatrix} : y_1, \dots, y_{n-1} \in \mathbb{R} \right\}$ has

dimension $n-1$.

This leads to the fact that θ^k may not converge a.e.